means for accessing a database structure providing a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database;

means for receiving in computer-readable form a text document to be classified;

processor means operable to compare terms appearing in the text document with the terms in the classified vocabulary and to determine from the comparison the category for the document; and

means for supplying a signal carrying data representing the text document and data associating the text document with the determined category.

4. (Amended) The apparatus according to claim 1, wherein the processor means is operable to determine the category for the document by determining from the comparison the category or categories of terms in the document, assigning weightings to the determined categories for the terms, and assigning the document being classified to the category having the highest weighting.

7. (Amended) The apparatus according to claim 4, wherein the processor means is operable, for each term in the classified vocabulary and in the text document, to share a predetermined weighting factor between each category associated with the term.

8. (Amended) The apparatus according to claim 1, wherein the accessing means is arranged to access a plurality of collocations also forming part of the database, each collocation being associated with a specific different one on the subject matter categories and each collocation including a plurality of terms exemplifying the associated category.

12. (Amended) A computer processing apparatus for classifying a document, comprising:

-2-

means for accessing a database having a database structure providing a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database and the database also containing a plurality of collocations each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category;

means for receiving in computer-readable form a text document to be classified;

processor means operable to compare terms appearing in the text document with the collocations to determine the collocation having the most terms in common with the document, and to allocate the category of the determined collocation to the document; and

means for supplying a signal carrying data representing the text document and data associating the text document with the determined category.

19. (Amended) The apparatus according to claim 7, wherein the accessing means is arranged to access the collocations from store means separate from the remainder of the database.

20. (Amended) The apparatus according to claim 1, further comprising store means configured to store the database.

21. (Amended) The apparatus according to claim 1, further comprising store means storing the database.

22. (Amended) The apparatus according to claim 1, wherein the database structure provides said plurality of subject matter categories as a tree structure including a plurality of main subject matter areas each divided into two or more subject matter areas.

23. (Amended) The apparatus according to claim 1, wherein the database structure provides said plurality of subject matter categories such that each category is defined by a subject matter area and a species or genus.

24. (Amended) The apparatus according to claim 23, wherein the database provides said plurality of subject matter categories such that the species or geni are people, places, organizations, products and technology.

25. (Amended) The apparatus according to claim 23, wherein the database structure provides said plurality of subject matter categories such that the species or genus are the same for each subject matter area.

26. (Amended) The apparatus according to claim 1, wherein the database provides categories in each of the following subject matter areas: the universe, the earth, the environment, natural history, humanity, recreation, society, the mind and human history.

27. (Amended) The apparatus according to claim 1, wherein the database structure is such that, for a given meaning, a term is associated with only one category and different meanings of the same term are associated with different categories.

28. (Amended) The apparatus according to claim 1, wherein the supplying means comprises means for storing a signal supplied by the supplying means on a computer readable medium.

29. (Amended) The apparatus according to claim 1, wherein the supplying means comprises means for forwarding a signal supplied by the supplying means to another processing apparatus.

30. (Amended) The apparatus according to claim 1, wherein the supplying means comprises means for displaying the information to a user.

31. (Amended) In a computer processing apparatus having means for accessing a database having a database structure providing a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database and means for receiving in computer-readable form a text document to be classified, a method of classifying documents comprising:

comparing terms appearing in the text document with the terms in the classified vocabulary;

determining from the comparison the category for the document; and

supplying a signal carrying data representing the text document and data associating the text document with the determining category.

34. (Amended) The method according to claim 31, further comprising determining the category for the document by determining from the comparison the category or categories of the terms in the document, assigning weightings to the determined categories for the terms, and assigning the document being classified to the category having the highest weighting.

35. (Amended) The method according to claim 34, further comprising assigning weighting by, for each term in the classified vocabulary and in the text document, sharing a predetermined weighting factor between each category associated with the term.

36. (Amended) The method according to claim 31, further comprising accessing a plurality of collocations also forming part of the database, each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category.

SUB B97

A7

39. (Amended) In a computer processing apparatus having means for accessing a database having a database structure providing a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database and the database also containing a plurality of collocations each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category and having means for receiving in computer-readable form a text document to be classified, a method of classifying documents comprising:

comparing terms appearing the text document with the collocations to determine the collocation having the most terms in common with the document;

allocating the category of the determined collocation to the document; and

supplying a signal carrying data representing the text document and data associating the text document with the determined category.

A8   SUB B107

43. (Amended) The method according to claim 36, further comprising accessing the collocations from store means separate from the remainder of the database.

SUB B117

A9

50. (Amended) The method according to claim 31, further comprising carrying out the supplying by storing a signal on a computer-readable medium.

51. (Amended) The method according to claim 31, further comprising carrying out the supplying by forwarding a signal to another processing apparatus.

52. (Amended) The method according to claim 31, further comprising displaying the information to a user.

53. (Amended) A database for use with an apparatus in accordance with claim 1, the database having a database structure providing a plurality of different subject matter

-6-

categories, the database containing a classified vocabulary including terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database.

56. (Amended) A database for use with an apparatus in accordance with claim 12, the database having a database structure providing a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database and the database also containing a plurality of collocations each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associate category.

65. (Amended) An apparatus for classifying electronic documents, comprising:

storage means storing a classification scheme having a plurality of collocations each collocation being associated with a respective different subject matter area and containing a set of terms which exemplify that subject matter area;

means for comparing terms used in a document to be classified with the terms in said collocations;

means for allocating the document being classified to the one of said collocations which said comparing means identifies as having the most number of terms in common with the document being classified;

means for associating with the document being classified a code representing the subject matter area of the allocation collocation; and

means for storing the document together with the associated code.

Please add new Claims 79-80 as follows:

-7-

79. (New) A computer processing apparatus for classifying documents, the-apparatus comprising:

a database having a database structure defining a classification scheme for terms, the classification scheme having subject matter data defining main and subsidiary subject matter domains into which terms can be classified and genera data defining a predetermined number of genera to which terms can be allocated, the classification scheme being such that a term can be allocated to more than one subject matter domain but to only one genus so that each specific combination of subsidiary subject matter domain and genus defines a unique category,

the database also having classified vocabulary comprising a set of terms classified in accordance with the classification scheme such that each term is associated with category data identifying the corresponding category,

the database also including a classification scheme data set which includes a respective different classification scheme data set item associated with each category,

each classification scheme data set item comprising a collocation including a list of terms that may be used to describe the function, appearance or relationship with other objects of classified terms in that category or that may be used in relation to terms in that category;

a receiver operable to receive in computer-readable form a text document to be classified;

a processor configured to compare terms in the text document with terms in at least one of the classified vocabulary and the collocations to determine a category for the text document; and

-8-

a signal supplier configured to supply a signal carrying data representing the text document and data associating the text document with the determined category data.

80. (New) A method of classifying documents, comprising:

providing a classification scheme having subject matter data defining main and subsidiary subject matter domains into which terms can be classified and genera data defining a predetermined number of genera to which terms can be allocated, the classification scheme being such that a term can be allocated to more than one subject matter domain but to only one genus so that each specific combination of subsidiary subject matter domain and genus defines a unique category;

providing a classified vocabulary comprising a set of terms classified in accordance with the classification scheme such that each term in the classified vocabulary is associated with category data identifying the corresponding category;

providing a classification scheme data set which includes a respective different classification scheme data set item associated with each category with each classification scheme data set item comprising a collocation including a list of terms that may be used to describe the function, appearance or relationship with other objects of classified terms in that category or that may be used in relation to terms in that category;

receiving data representing a text document to be classified; and

comparing terms in the text document with terms in at least one of the classified vocabulary and the collocations to determine a category for the text document.